

Measuring risk with expectile based expected shortfall estimates

Master Thesis submitted to

Prof. Dr. rer. nat. Wolfgang Härdle

Institute for Statistics and Econometrics

IRTG 1792 - High Dimensional Non Stationary Time Series

Humboldt-Universität zu Berlin



by

Philipp Gschöpf
(524851)

in partial fulfillment of the requirements
for the degree of
Master of Science in Statistics

July 18, 2014

Abstract

The concept of expectiles is here used to estimate expected shortfall for risk management applications. The underlying idea that will be presented is to employ an asymmetric (expectile) estimator. This enables the usage of advanced quantile estimators and parameter smoothing without a breakdown of the model. This is contrary to historical expected shortfall estimation which frequently breaks down in small samples and reduces to the maximum sample value. It is further inquired whether this approach enables a more accurate estimation. The robustness as well as consistency properties of this estimator are critically surveyed. The usage of parameter smoothing is examined and a functional expectile estimator is proposed as an example of a practical application. It is found to be biased but highly more accurate. A new methodology for the judgment of expected shortfall estimator accuracy is proposed and it is formally shown that the proposed method outperforms competing models in terms of prediction accuracy.

Keywords:

Expectile, asymmetric regression, expected shortfall, conditional autoregressive expectile

JEL Classification: C22, C58, G17, G11, G12

Dedication

This thesis is dedicated to my parents without whom my academic career would not have been possible as well as my beloved fiancé Natalia whose kindness lifts my mood on cloudy days. Your love and support made it possible for me to overcome all hurdles in life.

Acknowledgements

Above all I wish to thank Wolfgang Karl Härdle and Andrija Mihoci who supported my research and guided me through my first journey in the academic world.

Also I would like to thank the whole team of the Ladislaus von Bortkiewicz chair of statistics for their kind support and the great working environment. Especially Petra Burdejova who brought to my attention the possibilities of functional expectile calculation. Additional special thanks goes to my IRTG 1792 co-students, I am looking forward to spend three great years together with you.

Last but not least I would like to thank for the financial support of the IRTG 1792 in the form of my qualifying stipend and my exquisite workspace and from the SFB 469 in the form of access to financial data and computer infrastructure.



Contents

1	Introduction	1
1.1	Small samples in risk management	2
2	Risk Measures	4
2.1	Overview	6
2.1.1	Value at risk	6
2.1.2	Expected shortfall	6
2.2	Properties of risk measures	7
2.2.1	Coherence	8
2.2.2	Convex risk measures	10
2.2.3	Law invariance	11
2.2.4	Comonotonic Additivity	12
2.2.5	Spectral risk measures	14
2.2.6	Elicitability	14
2.2.7	Robustness	15
3	M-Estimators	18

3.1	Quantiles	18
3.2	Expectiles	19
3.3	Relation between Quantiles and Expectiles	21
3.4	M-quantiles	22
4	Methodologies	24
4.1	Historical Simulation	24
4.2	Extreme Value Theory	25
4.2.1	Extreme value theory based quantile estimation	28
4.3	Expectile based expected shortfall estimation	29
4.4	An estimation method for smooth expectile curves	30
4.5	GARCH Models	32
5	Simulation Results	33
5.1	Static Samples	33
5.2	Time Dynamic Samples	35
5.3	Extensions	38
6	Empirical results	42
6.1	Data	42
6.2	Testing methodology	43
6.3	Risk management application	44
6.4	Results	45
6.4.1	A Mincer-Zarnowitz regression	49
6.5	Research Outlook: Larger samples	50

7	Summary	52
7.1	Further extensions	53

List of Figures

2.1	Example indicating challenges resulting from VaR	10
4.1	FTSE 100 1% expected shortfall	31
5.1	Time static example	33
5.2	Expected shortfall estimates for a Normal-Laplace mixture distribution with linear increasing contamination parameter ε	35
6.1	Realized FTSE 100 losses and corresponding 1%-expected short- fall estimators using a moving window size of $n_w = 200$ for adaptive estimation. The moving window expectile (blue), historical (red), EVT GARCH(1,1) (green) and smooth ex- pectile (black) estimators are displayed.	44
6.2	Conditional scaled differences (ν^*) of 5% expected shortfall estimators for the UBS stock returns. A moving window size n_w of 200 was employed.	45
6.3	Conditional scaled differences (ν^*) of 1% expected shortfall estimators for the UBS stock returns. A moving window size n_w of 200 was employed.	46
6.4	Test statistics ν_t^* (6.2) for testing the null of unbiasedness of four different expected shortfall estimators: historical (red), moving window expectile (blue), smooth expectile (black) and extreme value theory (green). Here we set $\theta=0.05$	47

6.5	Conditional scaled differences of 1% expected shortfall estimators (Cisco stock)	48
6.6	Conditional differences ν^* , for a smooth expectile estimator with $\theta = 5\%$ expected shortfall (Cisco stock). The smooth quantile function was calculated over the entire sample (6115 observations), whereas τ was calibrated using a moving window of size $n_w = 600$	51

List of Tables

2.1	Properties of risk measures as in Emmer et al. (2013), extended by selected properties	8
5.1	Mean absolute difference (MAD) of expected shortfall estimators under a Normal-Laplace mixture distribution with fixed mixture parameter ε	34
5.2	Mean absolute difference in contaminated setting with mixture parameter ε increasing linearly in the sample	37
5.3	Mean and variance for selected expected shortfall estimators and an underlying Normal-Laplace distribution where the mixture parameter ε was increased slowly over the sample	38
5.4	Mean absolute difference under a partially known set of parameters	39
5.5	Mean squared difference for a Normal-Laplace mixture distribution with fixed mixture parameter ε	41
5.6	Bias with an underlying Normal-Laplace mixture distribution with fixed mixture parameter ε	41
6.1	95% confidence intervals for ν^* with $\theta = 0.05$	46
6.2	ES test statistic confidence intervals (using EVT quantiles) . .	49
6.3	95% confidence intervals for the β parameters in regression (6.4)	50

Chapter 1

Introduction

Current risk measures have been found to be unable to capture risk appropriately (historical estimation) or only with low efficiency (extreme value theory) when the reliable underlying sample is small. Low samples occur frequently in practical applications as financial portfolios often include illiquid assets or the underlying market conditions are fast changing such as the occurrence of structural breaks in the variance that potentially render large parts of the sample unfit for ad-hoc risk measurement.

A broad inquiry into risk measures, reveals that the expected shortfall is in fact the only risk measure that possess a very desirable combination of properties, namely consistency and comonotonic additivity. A drawback for its application has been lack of elicibility, i.e. with current methods verification and comparison of competing estimation procedures can only be executed with respect to relative performance. This has been restricting the development of testing and backtesting procedures for expected shortfall.

A short overview of M-quantile theory demonstrates the flexibility of expectiles. Relations to the concepts of expected shortfall and quantiles reveal usefulness thereof. A conducted simulation study provides statistical evidence on the degree of robustness of the presented estimator. Expectiles are elicitable. This enables the judgment of the overall goodness of fit of the employed estimator and makes the expected shortfall estimator conditionally elicitable.

A drawback to the presented method is its dependency on the quantile structure of the underlying stochastic process. The relation between expectiles

and quantiles is investigated and several possible approaches and solutions are proposed and evaluated. Results indicate that the estimation of the distance from the expectile to the quantile with current methods may introduce a bias into the model. The evidence provided in this study suggests that this challenge could possibly be overcome by employing a more efficient small sample estimator of the quantile and its relation to the expectile.

The presented methodology is successfully applied in practice, in financial risk management. Its ability to utilize smooth quantile and expectile estimators provided an increase in goodness of fit over comparable models. Even though it was found to be biased it dominated competing models, as the largest deviation from the realized expected shortfall was smaller than for all other models in the samples analyzed in this study.

Finally, further evidence is shown that the bias problem can potentially be overcome in large samples.

1.1 Small samples in risk management

Financial portfolio's often suffer from a small size of reliable samples. This is due to two phenomena that are frequently encountered in financial risk management:

- Illiquidity
 - Due to unequal spacing of the realizations the portfolio sample rate is often scaled down. Common practice is then to calculate the portfolio returns on a daily basis, even if the portfolio contains assets quoted with high frequency.
 - Many portfolios contain infrequently traded assets such as stocks or bonds of small sized companies or companies from illiquid markets.
- Time changing underlying stochastic processes
 - A financial asset will have time changing properties. In times of a market crash the variance and kurtosis of an asset may increase dramatically. This makes portfolio realizations that have not been observed in the immediate past ill-fit for risk measurement and assessment.

Many practical risk management applications rely on accurate measurement of risk in illiquid markets. A common example is the margin calculation for a financial title that depends on a medium sized company which is held in "big chunks" by large investors and thus rarely traded. The development of reliable risk management tools is a centerpiece and of utmost importance for practical risk management. This creates many challenges. For example the 99% quantile of a sample of 50 returns is in fact not defined and often simply reduced to the largest value to ensure that the estimator is always defined. This may not even be done consciously by the risk manager as many computer algebra systems use this routine without alerting the user. This naturally creates bias towards underestimating the portfolio risk in small samples.

A recent study by Taylor (2008) emphasized the existence of a relationship between expectiles and expected shortfall. The aim of this study is to analyze the performance of this approach and possible extensions in an applied risk management setting with a focus on small sample properties.

Chapter 2

Risk Measures

The purpose of risk measures in the financial industry is to quantify the risk of a portfolio. The question how to measure risk is of utmost importance for the vast majority of market participants:

- Investors
Risk assessment is an integral part of investment decision. Most investors will generally not be risk neutral and the decision whether to buy an asset will depend on its return as well as on its riskiness.
- Risk managers
The choice of measures to assess risk will substantially influence decisions of risk managers such as how much capital to set aside to prepare for extreme market events.
- Clearing houses
The calculation of margin requirements for the clearing members depends on how risk is measured.
- Regulators
An important part of modern financial market regulation has become the determination of risk reserves that market participants will be required to hold.

For the purpose of measuring risk, the random portfolio loss for an Asset with realizations S_t in period $t + 1$, $t \in \{0, \dots, T\}$ is approximated as:

$$L_{t+1} = -\{\log(S_{t+1}) - \log(S_t)\} \quad (2.1)$$

See for example *RiskMetricsTM - Technical Document* (1994) or Embrechts et al. (2005b). Note that under the Basel accord calculations are to be executed under the assumption of an unchanged underlying portfolio. To simplify notation a portfolio consisting of one asset is assumed below without loss of generality.

Let (Ω, \mathcal{F}, P) be a probability space with the sample space of possible events Ω , sigma algebra \mathcal{F} on subsets of Ω and probability measure $P : \mathcal{F} \rightarrow [0, 1]$. Following the setup of Föllmer and Schied (2004) let $M(P)$ be the set of all absolutely continuous probability measures with respect to P on this probability space and \mathcal{Q} be a class of probability measures on Ω .

The formal definition of a risk measure is simply a mapping from Ω that contains all possible loss scenario - such that $L \in \Omega$ will hold - to the positive real numbers:

$$\rho : \Omega \rightarrow \mathbb{R}_+ \quad (2.2)$$

If the mapping fulfills the following three conditions it is referred to as a risk measure:

1. $\rho(0) = 0$
If the portfolio is empty then there should be no risk indicated. Mappings adhering this condition are called **Normalized**.
2. $\rho(L + a) = \rho(L) - a \forall a \in \mathbb{R}$
If a nonrandom scalar (i.e. a cash position) is added then the risk should decrease by exactly the amount added. Risk measures need to be **Translative** as this greatly simplifies the determination of the necessary funds necessary risk reserves to compensate for risky positions.
3. $\rho(L_1) > \rho(L_2) \forall L_1 < L_2 \in \Omega$
Monotonicity ensures that (absolute) higher losses lead to higher risk indications.

2.1 Overview

In this section the most commonly employed risk measures will be introduced. Historically the most popular risk measure was the standard deviation:

$$\sqrt{\text{Var}(L)} = \sqrt{\mathbb{E}[(L - \mu)^2]} \quad (2.3)$$

With μ denoting the arithmetic mean of L .

2.1.1 Value at risk

The Basel II Accord, on banking supervision (2014) proposed the preferred approach to measure market risk to be value at risk (VaR). Since then VaR estimation has become a crucial part of risk management and driven by the Basel regulation framework VaR became the most important measure for market risk.

VaR is the Loss that the portfolio does not exceed with $\theta\%$ probability. With the cumulative distribution function F_t of L_t this definition can be formally written as:

$$\text{VaR}_t = F_{t+1}^{-1}(\theta) = \inf\{L; F_{t+1}(L) \geq \theta\} = q_\theta \quad (2.4)$$

Obviously this is equivalent to the quantile of the portfolio loss. It measures the worst loss that is to be expected with the current set information with confidence level θ . Basel II states that market participants should use $\theta = 1\%$ and a forecast horizon of 10 days. At least one year of historical data should be used to estimate VaR.

2.1.2 Expected shortfall

In recent years VaR has been criticized for a number of reasons, which will be explained in the next section. Expected shortfall (ES) was proposed as

an alternative risk measure.

$$\text{ES} = \mathbb{E}(L|L \geq F^{-1}(\theta)) \quad (2.5)$$

For continuous distributions of the loss L , the $\theta\%$ expected shortfall is given as:

$$\text{ES}_\theta = \mathbb{E}[L|L \geq F^{-1}(\theta)] = \int_{F^{-1}(\theta)}^{\infty} \frac{f(L)}{1-\theta} dL \quad (2.6)$$

See for example Embrechts et al. (2005b). Note that this definition (with " $>$ " and not " \geq ") is also well defined if F^{-1} has a probability atom at θ , Rockafeller and Uryasev (2002).

In it's (2014) consultative document the Basel Committee on Banking Supervision recommends the usage of expected shortfall for risk management purposes. Risk management departments have generally not yet implemented expected shortfall as well as VaR. Finding new methods to estimate and back test expected shortfall poses many possibilities for applied research. The remainder of this chapter explains why expected shortfall was recommended in a policy shift by the Basel committee.

2.2 Properties of risk measures

There exists an array of propositions for desirable properties that a "good" risk measure should have. In the following the most important these properties will be explained. Table 2.1 gives an overview, considering selected measures.

Property	Variance	VaR	ES	Expectile
Coherence			✓	✓
Comonotonic Additivity		✓	✓	
Robust, Weak Topology		✓		
Robust, Wasserstein	✓	✓	✓	✓
Elicitability		✓		✓
Conditional Elicitability	✓	✓	✓	✓

Table 2.1: Properties of risk measures as in Emmer et al. (2013), extended by selected properties

2.2.1 Coherence

Artzner et al. (1998) propose four axioms of coherent risk measures:

4. Translation invariance, $\rho(L + a) = \rho(L) - a \forall a \in \mathbb{R}$, see 2. Translative
5. Monotonicity, $\rho(L_1) > \rho(L_2) \forall L_1 < L_2$, see 3. above.
6. $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$

Subadditivity ensures that diversification never leads to higher risks. This is a natural assumption as Artzner et al. argue that market participants could simply open several accounts at a clearing house (or broker) and split their positions to reduce their margin requirement if it would not hold.

7. $\rho(kL) = k\rho(L) \forall k \geq 0$

Due to Subadditivity $\rho(L_1 + (k - 1)L_1) \leq k\rho(L_1)$ must hold. But since an increased position size will make it harder to liquidate the position the risk of the position should at least not be decreasing with it's size.

Positive homogeneity ensures that this is not the case.

Not only are the requirements for coherence well in line with what market participants reasonably could refer to as a "minimum consensus" about what properties a "good" risk measure should fulfill at least. Also from a computational point of view it is very advantageous to have a risk measure that fulfills the coherence properties as a dual representation then exists and can

be expressed as:

$$\rho(L) = \sup_{Q \in \mathcal{Q}} \mathbf{E}(L) \quad (2.7)$$

For a set \mathcal{Q} of risk measures on Ω , the set of possible losses, Artzner et al. (1998) as well as Huber (1981).

VaR is not subadditive for every possible loss distribution, as shown (among others) in Embrechts et al. (2009). Unfortunately many portfolios include derivatives that depend nonlinearly on the assets value. In this case the lack of subadditivity can become a severe factor. A simple example displaying VaR's lack of subadditivity in a portfolio that includes digital options can be found in exercise 16.11 in Franke et al. (2010).

Whether VaR will be subadditive actually depends on the distribution of the portfolio loss. Examples where VaR will be subadditive are:

- If the losses are identically, independently distributed (i.i.d.) and positively regularly varying, Embrechts et al. (2009)
- If the loss underlies an elliptical distribution, Embrechts et al. (2005b)

These results may lead to the fast - but wrong - conclusion that VaR should be "approximately subadditive" for real financial data as it's distribution could for example be well approximated by an elliptical distribution. However Embrechts et al. (2009) note that for skew, heavy-tailed risk distributions as they are common in financial applications this will not hold well. Especially the tails that are of most interest for risk management will not be well approximated.

An even more severe drawback for VaR is that it's calculation is mainly influenced by a small local area around the realized VaR. Initially seen as an advantage due to very high robustness this also means that the shape of the extreme tail does virtually have no impact on VaR. An example of a discrete distribution to display this is given in Figure 2.2.1 where the value of the leftmost possible realization does not have any impact on the 5% VaR. In this discrete setting it is obvious that the risk stemming from the extreme left tail obviously can change dramatically without much impact on VaR. If the CAPM holds, an asset with a heavier tail behind the considered

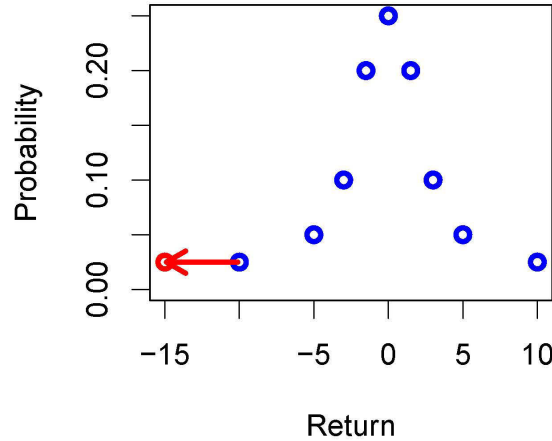


Figure 2.1: Example indicating challenges resulting from VaR

quantile will have higher mean returns but also higher risk. Thus VaR will most likely not give risk- and portfolio managers proper incentives.

In contrast to VaR expected shortfall is coherent, Delbaen (1998). And not only is it coherent but also it is the smallest consistent and law invariant risk measure that dominates VaR. Law invariant means that the risk of loss L depends only on the probability distribution F_L where $L \sim F_L$.

Additionally expected shortfall is dependent on the shape of the tail and thus less prone to the neglect of sudden tail-risk changes.

2.2.2 Convex risk measures

As indicated in the definition of positive homogeneity (see 7.) the definition of coherence by Artzner et al. (1998) poses problems for certain nonlinear increases in risk. Most importantly increased liquidity risks due to very large positions will not be considered appropriately. This prompted Föllmer and Schied (2002) to introduce the notion of convex risk measures. A risk measure is called a **convex risk measure** if it is:

8. Translation invariant, $\rho(L + a) = \rho(L) - a \forall a \in \mathbb{R}$, see 2.
9. Monotone, $\rho(L_1) > \rho(L_2) \forall L_1 < L_2$, see 3.
10. And **Convex**, i.e. $\rho(\lambda L_1 + (1 - \lambda)L_2) \leq \lambda \rho(L_1) + (1 - \lambda)\rho(L_2)$
 Note that this will still ensure that diversification never increases the portfolio's risk.

Another very convenient property that comes with convexity if $\rho(L)$ is normalized (see 1.) then, according to Föllmer and Schied, it "can be interpreted as a 'margin requirement', i.e., the minimal amount of capital which, if added to the position at the beginning of the given period and invested into a risk-free asset, makes the discounted position [...] 'acceptable'".

Like for coherent risk measures there exists a convenient explicit form of the dual optimization problem of the calculation of $\rho(L)$:

$$\rho(L) = \sup_{Q \in \mathcal{Q}} \{E_Q(L) - \theta(Q)\} \quad (2.8)$$

Where θ is a convex "penalty function", Föllmer and Schied (2002) and Heath (2000). In the dual representation it also becomes more obvious that the definition of coherent risk measures is more strict than that of convex risk measures. This can be easily verified by setting $\theta(Q) = 0$ - which is affine and thus convex. The dual representation of coherent risk measures is clearly a special case of convex risk measures. In this special case diversification is never increasing the risk but liquidity risk is not considered. Since coherent risk measures are a subset of convex risk measures it is not surprising that according to Föllmer and Schied (2008) any positive homogeneous and sub-additive risk measure is also convex. Expected shortfall and expectiles are both convex risk measures.

Obviously Value at risk fails to be a convex risk measure according to this definition as it is not subadditive and thus not convex.

2.2.3 Law invariance

As seen above it is often desirable to have a risk measure that depends only on the probability distribution of the losses of portfolio positions and not on

any other factors. Formally, if

$$\begin{aligned} \rho(L_1) &= \rho(L_2) \\ \text{with } P(L_1 \leq c) &= P(L_2 \leq c) \\ \forall c \in \mathbb{R} \end{aligned}$$

holds for two random then the risk measure is called **law-invariant**. Note that this definition does not rule out liquidity risks. Obviously VaR is law invariant.

It has been shown, Kusuoka (2001), Kunze (2003), Dana (2005), Föllmer and Schied (2004), Frittelli and Rosazza Gianin (2005) and for the non-continuous case Jouini et al. (2006), that every law-invariant risk measure can be represented as

$$\rho(L) = \sup_{\mu \in M} \int_{(0,1]} \text{ES}_{\theta}(x) \mu(d\theta) \quad (2.9)$$

where M is the set of all absolutely continuous risk measures. Also for every convex and law-invariant risk measure $\rho(L)$

$$\rho(L) \geq \text{ES}_{\theta}(L) \quad (2.10)$$

2.2.4 Comonotonic Additivity

L_1 and L_2 are comonotonic if there exist monotonic increasing functions $f_1(X)$ and $f_2(X)$ that map the same source of uncertainty X onto L_1 and L_2 :

$$\begin{aligned} L_1 &= f_1(x) \\ L_2 &= f_2(x) \end{aligned}$$

Thus comonotonicity between two random variables means that they are perfectly positive dependent. (Perfect negative dependence is defined only for

the two dimensional case and referred to as "countermonotonicity"). Speaking in a copula context the upper Frechet Hoeffding bound for cumulative distribution functions can only be achieved if two random variables are comonotonic.

A good way to imagine comonotonicity for financial instruments is that the losses occur only from the same source of uncertainty. A simple example is to look at two futures which only differ in their strike prices.

A risk measure $\rho(L)$ is **comonotonic additive** if for every comonotonic L_1 and L_2 it holds that:

$$\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$$

This is a highly desirable property. In practical risk management and clearing the calculation of the risk is commonly done in aggregated form. An example for this methodology is Eurex PRISMA, a cutting-edge risk management methodology used by Eurex. In PRISMA groups consisting of an asset and all derivatives that have this asset as underlying are formed prior to margin calculation. AG (2013) p. 3: "Cleared products that share similar risk characteristics will be assigned to so-called Liquidation Groups, which result in more accurate risk calculations and which will enable cross-margining within Liquidation Groups." Thus it is of crucial importance that risk measures consider aggregation effects properly, i.e. are comonotonic additive.

Comonotonic assets can be seen as a special case of assets that have the highest possible aggregated risk, Dhaene et al. (2002). This comes as no surprise as comonotonic assets will offer no diversification when they are aggregated. Lacking comonotonic additivity is a very serious drawback as it means that the risk measure is not considering aggregation and diversification effects properly.

Note that from the risk measures considered in this paper only Value at Risk (VaR) and Expected Shortfall (ES) are comonotonic additive risk measures, e. g. Emmer et al. (2013). Specifically note that expectiles are not comonotonic additive which marks the harshest argument against expectiles as risk measure.

2.2.5 Spectral risk measures

Acerbi (2002) introduces the class of spectral risk measures. Tasche (2002) shows (Theorem 3.6 and Remark 3.7 ii) that this class is simply the class of all risk measures that are

- coherent
- law invariant
- and comonotonoc additive

Tasche (2002) further shows (also in Theorem 3.6) that every coherent, law invariant and comonotonic additive risk measure will be of the following form:

$$\rho(L) = p \int_0^1 \text{VaR}_u(L) F(du) + (1 - p) \text{VaR}_1(L) \quad (2.11)$$

Where $p \in [0, 1]$. For $p = 1$ and $F(u) = \max(0, \frac{u-\tau}{1-\tau})$ expected shortfall is obtained.

VaR and expectiles are not spectral risk measures, Emmer et al. (2013). This comes quite naturally as Var is not coherent and expectiles are not comonotonic additive.

Unfortunately there is a conflict between subadditivity and robustness for spectral risk measures, Cont et al. (2010)

2.2.6 Elicitability

Gneiting (2011) and Ziegel (2013): A risk measure ρ is elicitable if it can be defined as the minimizer of a suitable expected scoring function. Elicitability is a very desirable property for computational efficiency, forecasting and testing algorithms. Gneiting (2011) gives a formal definition of elicibility. Let $S : \mathbb{R}^2 \rightarrow [0, \infty)$ be a scoring function. $\rho(L)$ is said to be elicitable with respect to a subset $M_1(P) \subset M(P)$ of all continuous probability measures iff

$$\rho(F_t(L)) = \arg \min_{\hat{L}_t} \int S(\hat{L}_t, L_t) dF_t(L) \quad \forall F_t \in M \quad (2.12)$$

Where \hat{L}_t are portfolio loss point forecasts and L_t are verifying observations. Expected shortfall is not elicitable, Gneiting (2011). Ziegel (2013) and Bellini and Bignozzi (2013) reveal that in fact expectiles are the only elicitable and coherent risk measure. But expected shortfall can be approximated by an average of quantiles, making it "conditionally elicitable", Tasche (2013). Additionally Taylor (2008) reveals a relation between expectiles - which are elicitable - and expected shortfall, see also equation (4.18). This relation can potentially be used to achieve conditional elicibility.

2.2.7 Robustness

According to Huber and Ronchetti (2009) "robustness signifies insensitivity to small deviations from the [model] assumptions". Naturally robustness against outliers is widely regarded as a desirable property for risk measures. But as already noted in the section above robustness can lead to trade-offs for risk measures.

This is exemplified by the Value at Risk - one of the most robust risk measures - of the distribution in fig. 2.2.1. It will be hard to react to sudden changes of the tail structure of the risk measure is too robust. In the given case the 10% VaR and 5% VaR are not reacting at all to the changed values of the 4% quantile because it mainly depends on a local area around the 10% quantile. Another more positive viewpoint of this is that the assumptions the model states on the tail behavior of the distribution have little impact on VaR.

A common measure for robustness is continuity with respect to the weak topology, see e.g. Huber and Ronchetti (2009). Stahl et al. (2012) argue that it is more appropriate to use continuity with respect to the Wasserstein metric instead. Most risk measures are robust in this sense, see table 2.1, especially the expected shortfall and expectiles are Lipschitz continuous with respect to the Wasserstein metric, Pflug and Wozabal (2007), Pichler ((forthcoming) or Bellini et al. (2014)

Robustness in an ε -environment

This survey will also consider a setup similar to Huber (1964). Consider a mixture distribution with parameter ε

$$F_\varepsilon = (1 - \varepsilon) N(0, 1) + \varepsilon H \quad (2.13)$$

where H is an unknown distribution. The minimax estimator θ_n is the estimator that has the lowest asymptotic variance

$$\min_{\theta_n} \left[\sup_{n \rightarrow \infty} \{\text{Var}(\theta_n)\} \right] \quad (2.14)$$

Consider now the class of estimators θ_n with loss function ρ defined by

$$\theta_n = \arg \min_{\theta} \sum_{i=1}^n \rho(Y_i - \theta) \quad (2.15)$$

The minimax estimator for 2.15 in 2.13 is, Huber (1964) and Huber and Ronchetti (2009)

$$\rho(u)_{opt} = \begin{cases} \frac{1}{2}u^2, & \text{if } |u| < k \\ -k|u| - \frac{1}{2}k^2, & \text{else} \end{cases} \quad (2.16)$$

Where $k \in [0, \infty)$ and decreasing in ε can be recovered from

$$\frac{2\varphi(k)}{k} - 2\Phi(-k) = \frac{\varepsilon}{1 - \varepsilon}$$

For a normal distribution the optimal estimator in the minimax sense would thus be the arithmetic mean. As the distribution gets contaminated (for increasing ε) the extreme values of the sample get limited to the cut off-value k . As the distribution becomes totally unknown this collapses to one point and the optimal estimate is the median. The least favorable density for this estimator, i.e. the worst possible distribution that has all contaminating mass outside the cut-off interval $[u - k, u + k]$ is

$$f_0(u_{opt}) = (1 - \varepsilon)(2\pi)^{-\frac{1}{2}} \exp\{-\rho(u)_{opt}\} \quad (2.17)$$

Inserting the minimax optimal estimator 2.16 into 2.17 delivers two interest-

ing special cases. For a standard normal distribution

$$f_0(u_{opt}) = (1 - \varepsilon)(2\pi)^{-\frac{1}{2}} \exp(-x^2) \quad (2.18)$$

which - as expected - is a standard normal distribution. As the distribution becomes totally contaminated ($\varepsilon \rightarrow 1$)

$$f_0(u_{opt}) \rightarrow (1 - \varepsilon)(2\pi)^{-\frac{1}{2}} \exp(-|x|) \quad (2.19)$$

which is in fact a Laplace distribution with infinite Kurtosis. The density of an asymmetric Laplace distribution is given by

$$f_{ALD}(x) = \theta(1 - \theta) \begin{cases} \exp\{-(1 - \theta)|x - \theta|\}, & \text{if } |x| < \theta \\ \exp(-\theta|x - \theta|), & \text{else} \end{cases} \quad (2.20)$$

Maximum likelihood yields

$$\frac{\partial L(x)}{x} = \frac{\partial - \log f_{ALD}}{x} = \begin{cases} (1 - \theta)|x - \theta|, & \text{if } |x| < \theta \\ -\theta|x - \theta|, & \text{else} \end{cases} \quad (2.21)$$

which will yield the quantile. A similar derivation for the asymmetric normal distribution yields the expectile.

To enable a Monte-Carlo simulation of the ε -environment, H was set parametrically to a Laplace distribution. For appropriate mixture parameters ($\approx \varepsilon \in [0.05, 0.4]$) This was found to generate artificial portfolio returns that generally are able to produce very similar sample features (Kurtosis etc.) to real financial data series.

Chapter 3

M-Estimators

3.1 Quantiles

The quantile of the loss L with underlying cumulative distribution function $F(l)$ is defined as

$$q(\theta)_t = F_t^{-1}(\theta) = \inf\{l; F_t(l) \geq \theta\} \quad (3.1)$$

Quantile regression (QR) has been introduced by Koenker and Bassett (1978). The underlying idea is to use an asymmetric absolute loss function to "punish" for residuals below ("on the left side") of the estimator differently from residuals above

$$Q(\beta, \theta) = \arg \min_{\theta} \sum_{i=1}^n \rho_{\theta}(\hat{Y}_i(\beta) - q_{\theta})$$

with

$$\rho_{\theta}(\lambda) = |\theta - \mathbf{I}(\lambda < 0)| |\lambda| \quad (3.2)$$

Analogously for the continuous case with probability density function $\frac{\partial F(l)}{\partial l} = f(l)$

$$\arg \min_{q_\theta} \left\{ (1 - \theta) \int_{-\infty}^{q_\theta} |u - q_\theta| f(u) du + \theta \int_{q_\theta}^{\infty} |u - q_\theta| f(u) du \right\}$$

Basset and Koenker show that the solution $q(\theta)$ to this linear convex optimization problem will yield a consistent estimator of the θ -quantile.

3.2 Expectiles

Newey and Powell (1987) extended the idea of Aigner et al. (1976) of a regression that uses a squared asymmetric check function

$$\rho_\tau(\lambda) = |\tau - \mathbf{I}(\lambda < 0)|\lambda^2$$

So that

$$e(L_i, \tau) = \arg \min_{e_\tau} \sum_{i=1}^n \rho_\tau(L_i - e_\tau)$$

Ordinary least squares (OLS) is a special case of asymmetric regression as can be verified by setting $\tau = 0.5$ where it also holds that $e(\tau)$ will be equal to the arithmetic mean of Y . Analog formulation for the continuous case holds

$$\arg \min_{e_\tau} \left\{ (1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau)^2 f(u) du + \tau \int_{e_\tau}^{\infty} (u - e_\tau)^2 f(u) du \right\}$$

Derivation holds with respect to e_τ holds

$$2(1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau) f(u) du + 2\tau \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du \quad (3.3)$$

$$\text{further } F(x) = \int_{-\infty}^x f(u) du$$

Newey and Powell coined the terms "asymmetric least squares" (ALS) and "expectile" for the resulting optimization solution e_τ . More recently the term

least asymmetrically weighted least square (LAWS) started to replace "ALS" to distinguish the abbreviation from alternating least squares (ALS).

Obviously one of the advantages of LAWS over quantile regression is that the resulting problem is a quadratic convex optimization problem, giving it an edge in terms of computational speed. The border of the possible solution space is having a continuous first derivative that is strictly monotone and convex around the global minimum and the solution is easily found by finding the point where the first derivative 3.3 is equal to zero

$$\begin{aligned}
(1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau) f(u) du + (1 - \tau) \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du = \\
(-\tau) \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du + (1 - \tau) \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du \\
(1 - \tau)(\mathbb{E}[X] - e_\tau) = (1 - 2\tau) \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du \\
e_\tau - \mathbb{E}[X] = \frac{(2\tau - 1)}{1 - \tau} \int_{e_\tau}^{\infty} (u - e_\tau) f(u) du \quad (3.4)
\end{aligned}$$

which is equivalent to equation 2.7 in Newey and Powell (1987). An implicit definition of the expectile can be derived by reformulation

$$e_\tau = \frac{(1 - \tau) \int_{-\infty}^{e_\tau} u f(u) du + \mu + \tau \int_{-\infty}^{e_\tau} u f(u) du}{(1 - \tau) F(e_\tau) + \tau (1 - F(e_\tau))} \quad (3.5)$$

The technical proof can be found in appendix A.

Expectiles are related to quantiles. A reformulation in the manner of Jones (1994a) shows in fact that expectiles are in fact quantiles of a distribution function $G(y)$ that is related to the "original" cumulative distribution function $F(y)$

Setting 3.3 equal to zero yields

$$\begin{aligned} & \tau \left(e_\tau - 2 \int_{-\infty}^{e_\tau} e_\tau f(u) du \right) + \int_{-\infty}^{e_\tau} e_\tau f(u) du = \\ & \tau \left(\int_{-\infty}^{\infty} u f(u) du - 2 \int_{-\infty}^{e_\tau} u f(u) du \right) + \int_{-\infty}^{e_\tau} u f(u) du \end{aligned}$$

$$\begin{aligned} & \tau \left\{ 2 \left(\int_{-\infty}^{e_\tau} u f(u) du - e_\tau \int_{-\infty}^{e_\tau} f(u) du \right) + e_\tau - \mu \right\} = \\ & \int_{-\infty}^{e_\tau} u f(u) du - \int_{-\infty}^{e_\tau} e_\tau f(u) du \end{aligned}$$

Defining the partial moment $P(Y) = \int^Y x f(x) dx$ holds

$$w(e_\tau) = \frac{P(e_\tau) - e_\tau F(e_\tau)}{2\{P(e_\tau) - e_\tau F(e_\tau)\} + e_\tau - \mu} \quad (3.6)$$

This is a density as follows from the more general proof following 3.14. As the 0.5 expectile is the arithmetic mean of the original function it will also hold that the median of $w(e_\tau)$ will coincide with the mean of F .

3.3 Relation between Quantiles and Expectiles

Pareto-like distributions with tail index β are defined as

$$F_x(x) = 1 - L(x)X^{-\beta}, \quad \text{with } \lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \forall t > 0 \quad (3.7)$$

Bellini et al. (2014) show that for Pareto-like distributions it holds that

$$\frac{\bar{F}\{e_\tau(X)\}}{\beta - 1} \approx 1 - \tau \quad \text{as } \tau \rightarrow 1 \quad (3.8)$$

Obviously it holds further

$$\frac{\bar{F}\{e_\tau(X)\}}{\bar{F}\{q_\tau(X)\}} = (\beta - 1) \frac{1 - \tau}{1 - \tau} \quad (3.9)$$

Establishing that for Pareto-like distributions the quantile will be larger than the expectile iff $\beta < 2$ and smaller than the expectile iff $\beta > 2$ for large values of τ . Thus for return distributions in the maximum domain of attraction (MDA) of the generalized extreme value distribution, i.e. if $\exists W$ that is generalized Pareto (GP) distributed with shape parameter γ , such that

$$\sup_{x \geq 0} |F_u(x) - W_{\gamma, \beta(u)}(x)| \rightarrow 0 \text{ for } u \rightarrow \infty \quad (3.10)$$

see theorem 18.9 in Franke et al. (2010). It has been found in previous studies that this assumption describes the tail structure of financial data sufficiently well to deliver competitive results for most applicational purposes in risk management, see e.g. Moscadelli (2004). In these Studies the realized value of β was found well above 2, being in line with the findings in this paper and the additional findings that expectiles were more mean centered than quantiles for the financial time series considered in this survey. Implications for applications in risk management are that the expectile will be a less conservative risk measure for typical financial data.

Further promising implications are a possible extension of the time dynamic expectile based expected shortfall estimator introduced later. Estimation of the tail parameter β could enrich the available information set, indicating the relative distance between the expectile and the quantile. This could potentially be used to eliminate the bias - found in section 6 below - in small sample expectile based expected shortfall estimators.

3.4 M-quantiles

For a loss sample $\{l_1, \dots, l_n\}$ Huber (1964) proposes M-estimators as location estimators of the form

$$\arg \min_T \sum_{i=1}^n \rho(l_i - T) \quad (3.11)$$

where the loss function ρ is convex and symmetric. Note that for $\rho = -\log(f)$ the maximum likelihood (ML) estimator results as a special case. The derivative of ρ is known as the influence function

$$\psi = \frac{\partial \rho}{\partial T}$$

This approach was extended by Breckling and Chambers (1988) into the class of M-quantiles. M-quantiles differ from M-estimators in allowing the loss function to be asymmetric. It is defined by setting the derivative of 3.12 equal to zero and adjusting ψ accordingly

$$\int \psi_p(x - \hat{\theta}) F(dx) = 0 \quad (3.12)$$

$$\psi_p(x) = \begin{cases} (1-p) & \text{if } (x < 0) \\ p\psi(x) & \text{else} \end{cases} \quad (3.13)$$

A proof analog to 3.6 for M-quantiles y yields

$$\frac{\int_{-\infty}^y \psi(x-y)f(x)dx}{2 \int_{-\infty}^y \psi(x-y)f(x)dx - \int \psi(x-y)f(x)dx} = p = w(x) \quad (3.14)$$

3.14 is a density, Jones (1994b)

- $\int_{-\infty}^y \psi(x-y)f(x)dx \xrightarrow{y \rightarrow -\infty} 0$ and then $w(x) \xrightarrow{y \rightarrow -\infty} 0$
- $w(x) \xrightarrow{y \rightarrow \infty} \frac{\int \psi(x-y)f(x)dx}{\int \psi(x-y)f(x)dx} = 1$
- Finally $\frac{\partial w(x)}{\partial x} \leq 0 \forall x$ follows from 2.2 in Jones (1994b)

Obviously least asymmetrically weighted squares (LAWS) and quantile regression (QR) are nested by the M-Quantile class of estimators.

Chapter 4

Methodologies

Due to the increased interest in expected shortfall a wide variety of estimation methods has been proposed recently. Considered in this study are:

- Historical estimation
- Extreme value theory
McNeil and Frey (2000)
- Expectile based
Taylor (2008)

Another possibility is saddlepoint approximation. This is generally used for a "first and quick" estimation and will in terms of accuracy often be dominated by other methods. This was also the case for the dataset employed in this study, which is why it will not be further considered below.

4.1 Historical Simulation

Kolmogorov's strong law of large numbers states that for an i.i.d. sequence Y_1, \dots, Y_n with $E(Y_i) = \mu < \infty$ and $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$

$$\bar{Y}_n \xrightarrow{a.s.} \mu \tag{4.1}$$

This also holds for the set of all L_j that exceed the quantile q_θ . Thus under the assumption of i.i.d. data, a simple approach to calculate the expected shortfall of L is the average of all values exceeding the empirical quantile during a limited period $[t - \Delta, t]$ where $\Delta \in \mathbb{N}$ is the size of the estimation window

$$\widehat{\text{ES}}_t^{\text{HIST}} = \sum_{i=t-\Delta}^t \mathbb{E}[L_i | L_i > q_\theta] \quad (4.2)$$

Historical Simulation is a nonparametric technique, even though i.i.d. structure is assumed. This yields the advantage that model misspecification errors are minimized. A full avoidance of model misspecification can not be achieved though as liquidity risk is not factored in due to the assumed constant portfolio structure, see e.g. Embrechts et al. (2005a)

4.2 Extreme Value Theory

Diebold et al. (2000) and McNeil and Frey (2000) proposed to use the peaks over threshold (POT) method for financial data. The following section is a short summary of the approach presented in McNeil and Frey (2000). Extreme value theory foundations are presented according to Franke et al. (2010), Embrechts et al. (1997) as well as Embrechts et al. (2005b).

The aim of this section is to introduce a semi-parametric approach to expected shortfall. The resulting estimator will estimate the quantile nonparametric and assume that a Generalized Pareto Distribution approximates the tail structure sufficiently well. The sample is assumed to be generated by a process of the form

$$L_t = \mu_t + \sigma_t z_t \quad (4.3)$$

Unfortunately Heteroscedasticity and autocorrelation are severe problems for classical extreme value theory due to the i.i.d. assumption that will be necessary for (4.6) below. The apparent conditional heteroscedasticity in financial data clearly violates this assumption. To make these problems less severe the data is "pre-whitened" with GARCH volatility processes. The approach proposed by McNeil and Frey (2000) is implemented by employing

an ARMA GARCH and then fitting a GPD to the residuals z_t . F_z denotes the unknown distribution of Z

Due to 4.3 expected shortfall prediction for period $t + 1$ can be executed by

$$\hat{\text{ES}}_{\theta}^{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{\text{ES}}^t \quad (4.4)$$

For any random variable X with distribution F it follows from Bayes theorem for the distribution of the excesses over a threshold u :

$$F_u(x) = P(L - u \leq l | L > u) = \frac{F(l + u) - F(u)}{1 - F(u)} \quad (4.5)$$

The generalized Pareto distribution (GPD) is defined as:

$$G_{\gamma, \beta}(l) = \begin{cases} 1 - \left(1 + \frac{\gamma l}{\beta}\right), & \text{if } \gamma \neq 0 \\ 1 - \exp\left(\frac{-l}{\beta}\right), & \text{else} \end{cases}$$

ξ and β are the shape and scale parameter of the GPD. The Pickands (1975) Balkema and de Haan (1974) theorem states that for an i.i.d. sequence (X_1, \dots) the threshold excess function F_u is well approximated for a wide class of distributions¹ by the GPD as long as u is "large":

$$F_u(l) \rightarrow G_{\gamma, \beta}(l), \text{ as } u \rightarrow \infty \quad (4.6)$$

The mean excess function is defined as:

$$e(u) = E(L - u | L > u) \quad (4.7)$$

For a generalized Pareto distribution with

$$\begin{aligned} 0 \leq u < \infty, & \quad \text{if } 0 \leq \xi < 1 \\ 0 \leq u \leq -\left(\frac{-\beta}{\xi}\right), & \quad \text{if } \xi < 0 \end{aligned}$$

¹Especially this holds for the maximum domain of attraction of the generalized extreme value distribution

the mean excess function can be calculated explicitly as

$$e(u) = \frac{\beta + \xi u}{1 - \xi} \quad (4.8)$$

For a positive loss sample L_1, \dots the mean excess function can be consistently estimated by

$$e_n(u) = \frac{\sum_{i=1}^n (L_i - u) \mathbf{I}_{L_i > u}}{\sum_{i=1}^n \mathbf{I}_{L_i > u}} \quad (4.9)$$

From (4.8) it becomes obvious that the GPD obviously is characterized by a linear mean excess function in the threshold u . Thus a proper u can be found by selecting a value that approximately linearizes the mean excess function. For this purpose the sample mean excess function can be plotted against the order statistic. Once this initial threshold u is found the mean excess function for any threshold $v > u$ can be obtained by

$$e(v) = \frac{\beta + \xi(v - u)}{1 - \xi} = \frac{\xi v}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}, \text{ if } \xi < 1 \quad (4.10)$$

Especially this holds for any quantile $q_\theta > u$ of the residual losses $-z$. For practical purposes $F(u)$ can be estimated by the sample ratio of exceedances over u : $\frac{N_u}{n}$. Obviously a sufficient amount of exceedances is needed to estimate reliably. For the case where insufficient data is available an estimator for $F(u)$ was proposed by Smith (1987).

Assuming the tail of Z follows a GPD, the expected shortfall is

$$\mathbf{E}(G|G > \gamma) = \frac{\gamma + \beta}{1 - \xi} \quad (4.11)$$

Then also hold

$$\begin{aligned} Z - q_\theta | Z > q_\theta &= (Z - u) - (q_\theta) | (Z - u) > (q_\theta) \\ Z - q_\theta | Z > q_\theta &\sim \text{GPD}\{\xi, \beta + \xi(q_\theta - u)\} \end{aligned} \quad (4.12)$$

with 4.12 and 4.10 the expected shortfall follows as

$$ES_\theta = q_\theta + e(q_\theta) = \frac{q_\theta}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi} \quad (4.13)$$

The choice of u is a drawback of the model. Several possible choices have been implemented and a strong dependency of the estimator on u was noted. The (arbitrary) proposal of McNeil and Frey (2000) to use $u = q_{0.1}$ delivered comparative estimation results in the simulation and empirical studies below.

4.2.1 Extreme value theory based quantile estimation

Smith (1987) provides an extreme value based quantile estimator. An obvious estimator for the tail distribution (of observations exceeding the threshold u)

$$1 - F(x) = \{1 - F(u)\}\{1 - F_u(x - u)\} \quad (4.14)$$

is acquired by estimating $F(u)$ by the in sample proportion of threshold exceedances $\frac{N_u}{N}$ and $\{1 - F_u(x - u)\}$ by an extreme value distribution

$$\widehat{F}(x) = 1 - \frac{N_u}{N} \left(1 + \widehat{\xi}_N \frac{x - u}{\widehat{\beta}_N} \right)^{\frac{-1}{\widehat{\xi}_N}} \quad (4.15)$$

Inversion yields

$$\widehat{q}_\theta = q_u + \frac{\widehat{\beta}_u}{\widehat{\xi}_u} \left\{ \left(\frac{1 - \theta}{\frac{N_u}{N}} \right)^{\widehat{\xi}_u} - 1 \right\} \quad (4.16)$$

To ensure that q_u is always well inside the sample, u is fixed and chosen such that q_u is significantly closer to the center of the distribution than q_θ . Loosely following McNeil and Frey $u = 10\%$ was found to perform well in moving window estimation sizes of $\{50, 100, 200\}$ with respect to estimation efficiency and linearity of the mean excess function. Thus $\frac{N_u}{N} = 0.1$

4.3 Expectile based expected shortfall estimation

The conditional autoregressive expectile (CARE) model was introduced by Taylor (2008). It is based on asymmetric regression techniques - Aigner et al. (1976) - and the Conditional Autoregressive Value at Risk (CAViaR) model by Engle (1982). The purpose of CARE is to forecast expected shortfall in a time dynamic setting. The following section will give a short introduction to these foundations and then propose an algorithm - based on the results in Taylor (2008) - to measure expected shortfall.

Reformulation of 3.4 holds a relationship between expectiles and the expected shortfall

$$\frac{1-2\tau}{\tau} \mathbb{E}[(Y - e_\tau) \mathbf{I}\{Y < e_\tau\}] = e_\tau - \mathbb{E}[Y]$$

For scalar $e(\tau)$ thus, Taylor (2008)

$$\mathbb{E}[Y|Y < e_\tau] = e_\tau + \frac{(e_\tau - \mathbb{E}[Y])\tau}{(1-2\tau)F(e_\tau)} \quad (4.17)$$

For an appropriate τ with $e_\tau = q_\theta$ holds

$$\text{ES}_\theta = e_\tau + \frac{(e_\tau - \mathbb{E}[Y])\tau}{(1-2\tau)\theta} \quad (4.18)$$

This relationship between expected shortfall expectiles can be used to measure and forecast expected shortfall. The approach employed in this survey follows the following algorithm:

1. Start with a moving window subsample L_1, \dots, L_{n_w} of size n_w of the available sample L_1, \dots, L_n to allow for time adaptive estimation
2. Estimate the sample quantile \hat{q}_θ . This can be done historically with a pre-smoothed density function. Especially in small samples pre-smoothing led to increased prediction power.

3. With the sample expectile e_τ find τ by

$$\arg \min_{\tau} \{\hat{q}_\theta - \hat{e}_\tau\}$$

4. Relation 4.17 yields $\widehat{\text{ES}}^{EXPECTILE}$
5. Shift the moving window subsample to include L_2, \dots, L_{n_w+1}

q_θ was estimated within the moving window sample via historical estimation unless otherwise noted. In extremely small samples this could alternatively be done with quantile regression or local adaptive. The method has the additional property of producing quite similar results to the historical estimation, which allows for a good comparison of the two methods.

4.4 An estimation method for smooth expectile curves

This section will give a short insight into an alternative estimation method for the expectile function. This is considered due to an increase in performance of expectile based expected shortfall estimation when smoothing was employed.

Schnabel and Eilers (2009) propose a functional quantile estimator. The method is well implemented in the R package Expectreg, Sobotka et al. (2011). It is part of the library expectreg. The approach was found to deliver very competitive results in terms of goodness of fit. Let c_k , $k = 1, \dots, K$ be a set of scalars and ζ_k a set of appropriate basis functions. In the empirical application a basis of parabolas was found to perform very well. A function $x(t)$ can then be approximated by

$$x(t) \approx \sum_{k=1}^K c_k \zeta_k(t) \quad (4.19)$$

Figure 4.1 displays the 1% expected shortfall of the FTSE 100 estimated using the Schnabel and Eilers (2009) approach with Schall smoothing algorithm, Schall (1991). Apparently the increase in risk in the years 2007 and 2008 seems to be captured more appropriately by this more sophisticated

technique than with the simple moving window. This will be confirmed by a new measurement accuracy test for expected shortfall in the empirical section.

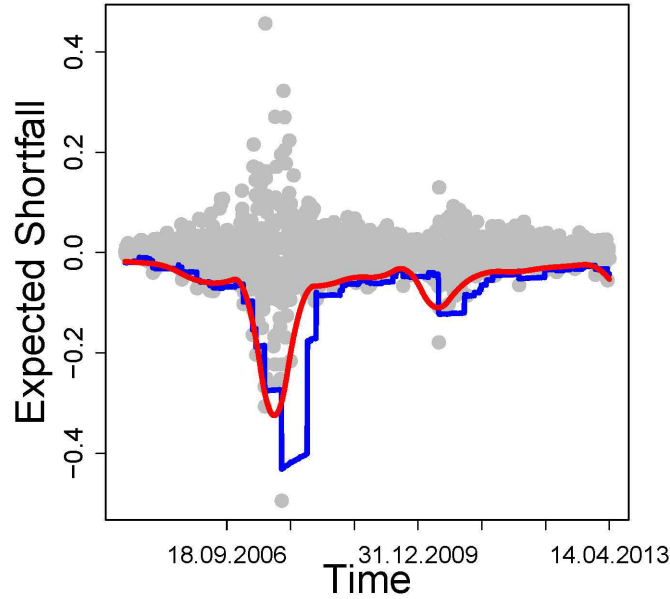


Figure 4.1: FTSE 100 1% expected shortfall

To allow analysis of its small sample properties, the approach was implemented using the same moving window sample of size n_w as for the other approaches above. The last value of the moving window expectile function was then taken as expectile estimator. This will generally produce a less smooth expected shortfall estimator than in figure 4.1, which in fact is a large sample version of the estimator proposed in this paper. Below it will be shown, that the small sample version of the estimator is prone to underestimate the risk with a biased estimation. This problem can be fixed however, if the Schnabel and Eilers (2009) expectile estimator without moving window as in 4.1 is employed.

In the following the rolling window approach employing the Schnabel and Eilers (2009) expectile estimator will be referred to as "Smooth expectile" approach.

4.5 GARCH Models

Pre-whitening can often increase the performance of expected shortfall estimation. For this purpose the volatility of the time series is estimated with a GARCH type model and the losses are regularized

$$L^* = \frac{L}{\hat{\sigma}_t} \quad (4.20)$$

The ARCH model was introduced by Engle (1982). Bollerslev (1986) expands the ARCH model into the GARCH model. Characteristics that occur with financial data often include heteroscedasticity. In the GARCH model it is assumed, that the data can be modeled by utilizing a ARMA process and the variance of the distortion parameter can be modeled by a second ARMA process. This enables enhanced modeling of volatility clustering and underlying leptocurtic distributions. The GARCH structure is given by

$$\begin{aligned} L_t &= \mu + \beta_1 L_{t-1} + \beta_2 L_{t-2} + \dots + \sigma_t \varepsilon_t + \gamma_1 \sigma_{t-1} \varepsilon_{t-1} + \gamma_2 \sigma_{t-2} \varepsilon_{t-2} \dots \\ \sigma_t^2 &= a_0 + a_1 L_{t-1}^2 + a_2 L_{t-2}^2 \dots + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-2}^2 + \dots \end{aligned}$$

The methodologies have been implemented in a rolling time window estimation procedure. As the aim is to find expected shortfall measures that perform good in small samples.

Chapter 5

Simulation Results

The properties of the proposed expected shortfall estimators will now be examined and compared in a controlled Monte Carlo simulation environment.

5.1 Static Samples

Let $F_\varepsilon = (1 - \varepsilon)N + \varepsilon H$ be a normal distribution contaminated by a distribution H . This setup follows Huber (1964). In the following H will be assumed to follow a Laplace distribution. Increasing ε will thus increase the kurtosis of this mixture. In this section ε will be fixed over the sample.

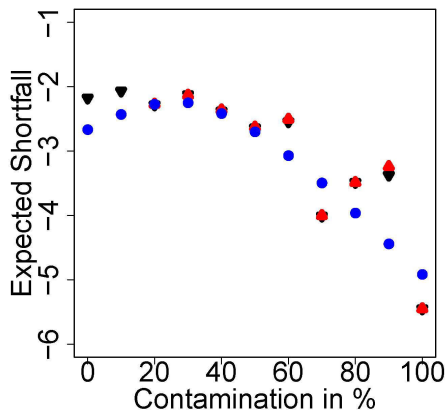


Figure 5.1 displays the true 1% expected shortfall of F_ε for 11 samples with fixed $\varepsilon \in \{0, 0.1, \dots, 0.9, 1\}$ in blue. The values were calculated as historical expected shortfall in a Monte Carlo simulation with sample size 1,000,000. This delivered in four different iterations results that had differences within $\mathcal{O}(10^{-5})$, indicating sufficient approximation of the true value.

Figure 5.1: Time static example A second sample with size $n = 100$ was generated. The expectile based expected shortfall estimator (black) was

computed using the historical quantile, smoothed with a normal kernel. Red triangles display the historical expected estimates. The most striking advantage of the expectile based method is that it is sure to deliver results whereas the historical method can break down due to the low sample size.

With respect to accuracy the expectile based approach delivers comparable results to the historical approach. 10,000 simulations with sample size 150 (to make sure the historical approach does not break down) were executed to calculate the mean absolute deviation (MAD) of the 1% expected shortfall estimators, see table 5.1. A slight superiority with respect to the MAD of the expectile based approach was only found employing smoothed underlying quantiles, reflecting a better quantile estimation.

n	10		150		
ε	Expectile	Historical	Expectile	Historical	Difference %
0	0.293	0.309	0.953	1.148	0.829
0.1	0.267	0.281	0.872	1.056	0.825
0.2	0.254	0.267	0.821	0.996	0.824
0.3	0.279	0.287	0.852	1.017	0.837
0.4	0.348	0.356	0.976	1.168	0.836
0.5	0.435	0.443	1.197	1.385	0.864
0.6	0.534	0.541	1.425	1.635	0.871
0.7	0.611	0.616	1.665	1.916	0.869
0.8	0.708	0.721	1.908	2.193	0.869
0.9	0.810	0.820	2.162	2.479	0.872
1	0.894	0.903	2.377	2.754	0.863

Table 5.1: Mean absolute difference (MAD) of expected shortfall estimators under a Normal-Laplace mixture distribution with fixed mixture parameter ε

The last two columns of table 5.1 show the MAD for a smaller sample size of $n=10$. For the historical method the largest value was used as it would otherwise break down and the quantile was not presmoothed, as this was introducing a bias due to the small sample size. The last row displays $\frac{\text{MAD}_{\text{Expectile}}}{\text{MAD}_{\text{Historical}}}$. Apparently the expectile based approach dominates even more for the leptokurtic Laplace distribution in this small sample setting.

These results show a slight superiority due to the possibility to use quantile

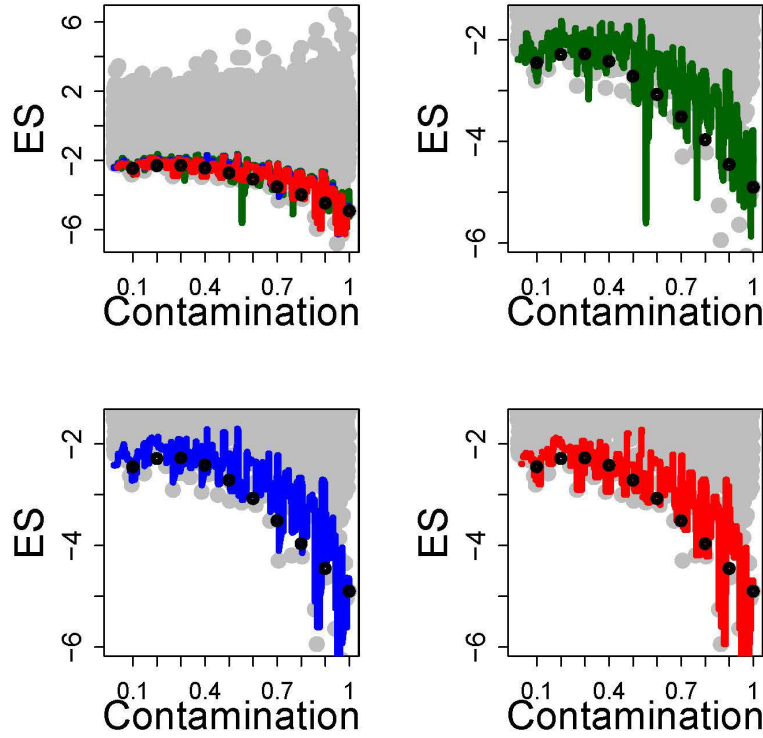


Figure 5.2: Expected shortfall estimates for a Normal-Laplace mixture distribution with linear increasing contamination parameter ε

 ESdynamicSIMULATIONTest.R

smoothing in the expectile based approach. This simple model is constructed to deliver results close to historical estimation.

5.2 Time Dynamic Samples

In the previous section a stationary setting with a sample generated by a distribution $F_\varepsilon = (1 - \varepsilon)N + \varepsilon H$ with fixed ε was analyzed. This section examines the properties of the proposed expected shortfall estimators under a time changing distribution. This is achieved by generating a sample of losses, starting with $L_1 \sim F_0$ and then gradually increasing $\varepsilon \in [0, 1]$ by a fixed step-size.

Figure 5.2 displays an example realization of this process where ε was increased in 0.0001 increments. The resulting sample size is $0.0001^{-1} = 10,000$.

The true expected value for selected points of the distribution was calculated via historical estimation using a simulated sample with 1 million realizations. In figure 5.2 the true values are displayed as black dots. The expected shortfall estimates were calculated using a rolling window size of 200. Historical estimation is displayed in red, expectile based in blue and extreme value theory based in green. The first 200 sample observations were used to initialize the model parameters and are afterwards discarded. As in the stationary case it was found that pre-smoothing with a Gaussian kernel increased the prediction power of the expectile based approach (as measured by the mean absolute deviation from the true value).

Table 5.2 gives the mean absolute deviation from the true value for various increment step sizes for ε and a moving estimation window size of 50. The average was calculated using 100 iterations of the algorithm. Important facts are observable in table 5.2

- The expectile based expected shortfall estimate shows similar estimation power as the historical estimate. The mean average distance is very similar for both models. This is a confirmation that the algorithm is implemented properly.
- No significant additional improvement of the expectile based approach over the historical approach when the underlying distribution of the sample changes faster (for larger incrementation steps of ε) can be found. This is of no surprise as this effect will mainly arise in prediction contexts and not in same period measuring.

For a non smoothed quantile function the expectile based approach was again found to be similar to the historical approach. As in the static case it is observed that a advantage of the expectile based approach over historical estimation in measuring is the fact that it allows for smoothing of the quantile function. This is due to the fact that the historical estimate is not defined anymore if the smoothed quantile exceeds the largest value in the estimation sample. Obviously further research opportunities arise in finding an optimal kernel to pre-smooth the quantile function. As is obvious from table 5.2 mentioned above the usage of a normal kernel increased the predictability of the expectile based approach but has very limited improvement potential for extreme quantiles. Quantile regression, e.g. Koenker (2005) could pose another possible solution to this challenge.

ε increments		0.0001			0.001		
ε	Expectile	Historical	EVT	Expectile	Historical	EVT	
0.1	0.493	0.498	0.548	0.455	0.459	1.354	
0.2	0.416	0.420	0.515	0.552	0.557	0.661	
0.3	0.450	0.456	0.545	0.454	0.460	0.528	
0.4	0.532	0.537	1.188	0.542	0.547	0.845	
0.5	0.702	0.705	0.935	0.733	0.739	1.042	
0.6	0.904	0.909	1.359	0.844	0.851	1.575	
0.7	0.971	0.978	1.418	1.114	1.120	1.773	
0.8	1.290	1.298	1.345	1.211	1.218	2.474	
0.9	1.475	1.483	1.834	1.251	1.256	4.435	
1	1.398	1.406	3.205	1.356	1.364	2.139	

Table 5.2: Mean absolute difference in contaminated setting with mixture parameter ε increasing linearly in the sample

 ESdynamicSIMULATIONTest.R

The controlled simulation environment allows for construction of tests. The true value of the expected shortfall is calculated as above with 1 million iterations. Next the variance of the estimators is approximated as squared difference to mean estimated value in all selected points over 500 simulation iterations. For a setting with ε incrementation steps of 0.001 and a moving estimation window of 50 the approximated mean and variance are given in table 5.3

ε	Mean				Variance	
	True ES	Expectile	Historical	EVT	Historical	Expectile
0.1	-2.431	-2.102	-2.092	-2.391	0.211	0.213
0.2	-2.277	-1.950	-1.941	-1.947	0.144	0.145
0.3	-2.267	-1.877	-1.869	-1.595	0.185	0.187
0.4	-2.408	-1.955	-1.946	-1.670	0.223	0.225
0.5	-2.702	-2.093	-2.083	-1.871	0.376	0.380
0.6	-3.083	-2.389	-2.378	-2.714	0.495	0.500
0.7	-3.505	-2.698	-2.686	-5.717	0.760	0.768
0.8	-3.956	-2.955	-2.941	-2.980	0.918	0.927
0.9	-4.428	-3.310	-3.295	-3.505	1.464	1.478
1	-4.910	-3.821	-3.803	-3.640	1.326	1.340

Table 5.3: Mean and variance for selected expected shortfall estimators and an underlying Normal-Laplace distribution where the mixture parameter ε was increased slowly over the sample

 `ESdynamicSIMULATION.R` (uncomment the squared differences first)

Due to the small sample size extreme value theory did not produce consistent variance estimators. In fact it also failed to consistently produce plausible estimates as extreme exceedance estimates were observed. This occurred rather rare (during the calculation of the values in table in 5.3 0.228833% of all estimates exceeded an MAD of 5) but indicates analogous problems to historical estimation in small samples. Expectile based estimation was in fact the only technique that was sure to deliver estimates in this survey. Historical estimates could only be obtained by substituting the expected shortfall with the maximum value of the sample as soon as θ^{-1} exceeded the sample size n .

5.3 Extensions

Further extensions of the expectile estimation or quantile estimation methods allows further improvement of the relative performance of the expectile based approach compared to historical estimation.

Table 5.4 displays the mean absolute distance $MAD_{Expectile}$ from the estimated 1% expected shortfall to the true value in a static setting. Three scenarios with partially known parameters are considered to infer on param-

eter influence. Either τ or $e(\tau)$ are unknown, as a reference the case where both are unknown is also given. The values displayed have been averaged over 100.000 iterations of the simulation, n was set to 100. The scenario with known e_τ clearly has the smallest MAD, emphasizing the need for an appropriate expectile estimator.

ε	Both unknown	τ known	e_τ known
0	0.3588234	0.3157145	0.2516114
1	1.065401	0.897209	0.745054

Table 5.4: Mean absolute difference under a partially known set of parameters

 ESdynamicSIMULATION.R

Among many possible expectile estimation methods, the algorithm proposed in Schnabel and Eilers (2009) was found to deliver results that dominated the reference methods in all considered empirical samples in terms of goodness of fit. Like the other methods it was implemented in a moving window, to allow analysis of it's small sample properties. Estimation fit relative to the moving window method was found to increase for larger samples This was to be expected as the approach delivers a time adaptive expectile fit.

Recall Figure 4.1, that among other desirable properties demonstrated the possible usage of the Schnabel and Eilers (2009) expectile estimation method with enhanced integration of larger samples into the information set. The moving window estimator on the other hand primarily works with small local samples to ensure adaptiveness. Note however that for large samples τ should still be estimated adaptively, as will be shown below.

In the following the smooth expectile approach is implemented using the (same moving window) sample as for the historical estimation. This is only done because it allows for comparison of the small sample properties that are of special interest. The "smooth" expectile approach will then actually look rugged, like in figure 6.1

Table 5.5 displays the mean squared difference of three different 1% expected shortfall estimators towards the true expected shortfall (calculated in a sample with 1,000,000 observations) in a static contaminated setting. The values were calculated employing a normal kernel density estimator for pre-smoothing the data before quantile computation. 1000 iterations of the algorithm were executed to ensure robust results. The available sample size

for expected shortfall estimation was set to 100.

The approach employing the smooth expectile function performs poor compared to the moving window expectile approach. This reflects the fact, that the moving window approach assumes (local) stationarity within the window. As this assumption is valid in this stationary simulation it dominates the approach employing a smooth expectile function. In practical applications however this assumption does not hold and the smooth functional approach was found to clearly dominate.

Another important characteristic of table 5.5 is that the extreme value approach tends to produce inconsistent results as ε increases. This is due to the artificial threshold u that was set at the 90% quantile. This introduces a form of expert bias into the model. As the kurtosis of the underlying model increases it becomes more likely that the threshold is inappropriately set. This can lead to unrealistic predictions. These occur rather rarely and are obvious but poses challenges in the automation of the extreme value theory approach in practical implementations.

Table 5.6 displays the estimation bias in the same setting as for table 5.5. It was approximated as the mean difference towards the true value $\widehat{\text{ES}} - \text{ES}_{\text{TRUE}}$. Apparently all methods underestimate the portfolio risk due to the small sample size. The smoothing of the expectile function was found to increase the bias, a common problem of smoothing techniques. Nevertheless it was found to dominate the other models in terms of prediction accuracy for real financial time series (formal tests in the following section).

ε	Expectile	Smooth expectile	EVT
0	0.210	0.476	0.717
0.1	0.179	0.392	1.107
0.2	0.157	0.347	0.458
0.3	0.194	0.406	9.504
0.4	0.295	0.534	16.710
0.5	0.472	0.888	1.704
0.6	0.689	1.114	4.510
0.7	0.894	1.585	9.986
0.8	1.180	2.212	24.250
0.9	1.398	2.473	1378.25
1	1.815	3.107	50.092

Table 5.5: Mean squared difference for a Normal-Laplace mixture distribution with fixed mixture parameter ε

 ESsmoothSIMULATION.R

ε	Expectile	Smooth expectile	EVT
0	0.124	0.556	0.248
0.1	0.162	0.524	0.258
0.2	0.097	0.473	0.160
0.3	0.171	0.527	0.152
0.4	0.190	0.602	0.391
0.5	0.180	0.703	0.295
0.6	0.252	0.864	0.331
0.7	0.293	1.029	0.317
0.8	0.343	1.145	0.288
0.9	0.357	1.302	1.619
1	0.483	1.496	0.939

Table 5.6: Bias with an underlying Normal-Laplace mixture distribution with fixed mixture parameter ε

 ESsmoothSIMULATION.R

Chapter 6

Empirical results

The calculation of risk measures for the current portfolio market exposure is the basis for margin calculations, e.g. Risk Based Margining, AG (2007). The focus of this section is to determine the goodness of fit of the expected shortfall estimation techniques introduced above. For this purpose a wide selection of financial time series was employed. The most striking results will be introduced in this section. This is supported by a selected set of examples, the results hold true for all time series covered in the full study.

6.1 Data

The data employed was taken from a Bloomberg terminal and Reuters datstream with kind permission of the research data center at the SFB 649. The dividend adjusted log-returns of several financial time series were employed in the survey

- FTSE 100 realizations from 01/09/1997 to 02/05/2005, GBP
- Stock, Cisco Systems, Inc. from 22/01/2007 to 18/02/2014, USD
- Stock, UBS AG from 21/01/2007 to 19/02/2014, EUR
- Stock, City bank Ltd. from 02/01/1990 to 07/04/2014, USD
- Stock, British petrol p.l.c. from 01/01/1971 to 26/05/2014, GBP

- Future (current month delivery), Brent crude oil from 01/01/1990 to 26/05/2014, USD

6.2 Testing methodology

McNeil and Frey (2000) propose a testing methodology for time dynamic expected shortfall estimates. This approach poses the same assumptions as the extreme value theory based estimator introduced above. Especially 4.3 holds.

Let the residual of the realized return r_t , conditional on the exceedance of the θ quantile $q_{t,\theta}$ be denoted by

$$\nu_t = \left\{ r_t - \text{ES}_\theta^t \mid \frac{r_t}{\sigma_t} > q_\theta \right\} \quad (6.1)$$

To have a lower dependence on the employed quantile estimation method the condition $r_t > q_{t,\theta}$ was evaluated using pre-whitened data. Under 4.3 these conditional residuals will be i.i.d. with $\mathbb{E}[\nu_t] = 0$. Standardization yields

$$\frac{\nu_t}{\sqrt{\text{Var}(r_t \mid \frac{r_t}{\sigma_t} > q_\theta)}} = \nu_t^* \sim (0, 1) \quad (6.2)$$

A bootstrap test, see e.g. Efron and Tibshirani (1993), is implemented to estimate $\text{Var}(Z \mid Z > q_{t,\theta})$. Of interest will be mainly a test in the fashion of

$$H_0 : \mathbb{E}[\nu_t^*] = 0 \quad H_1 : \mathbb{E}[\nu_t^*] \neq 0 \quad (6.3)$$

6.3 Risk management application

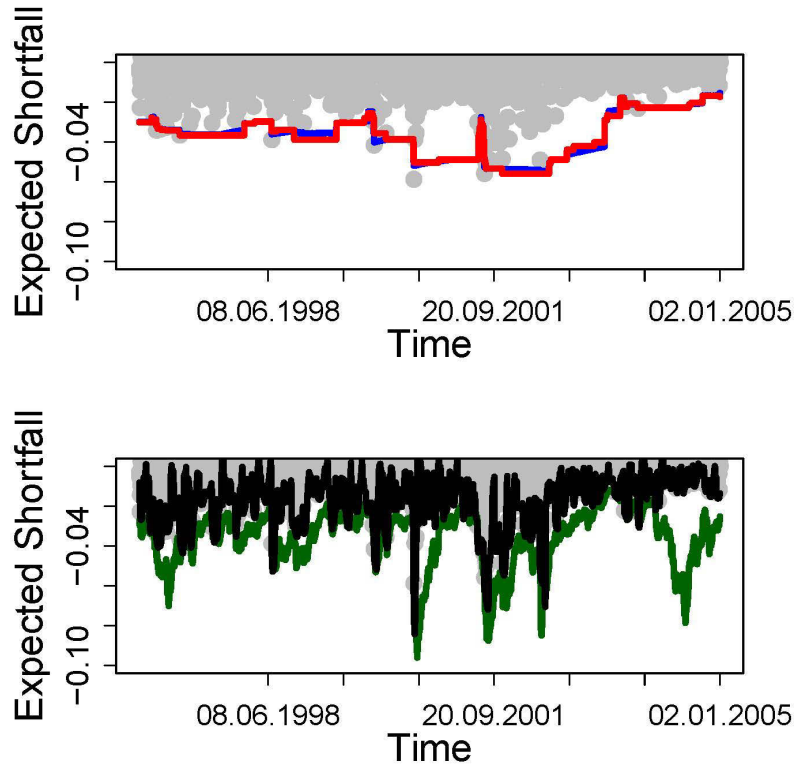



Figure 6.1: Realized FTSE 100 losses and corresponding 1%-expected shortfall estimators using a moving window size of $n_w = 200$ for adaptive estimation. The moving window expectile (blue), historical (red), EVT GARCH(1,1) (green) and smooth expectile (black) estimators are displayed.  ESPlot.R

Figure 6.1 displays expected shortfall estimators, estimated with a moving window size n_w of 200. Apparently the extreme value theory based approach indicates higher portfolio tail-risk than the other estimation approaches. Below it will be shown formally that actually only the extreme value theory based approach delivers unbiased estimation results. The smooth expectile approach shows a more volatile estimate, reflecting the fact that it captures the risk more adaptively than the simple moving window approaches (which are displayed in the upper graph).

6.4 Results

To judge the goodness of fit of the proposed expected shortfall estimator an approach following McNeil and Frey (2000) is implemented. A GARCH(1,1) model was employed for standardization and rescaling. Confidence intervals are calculated using a nonparametric bootstrap following Efron and Tibshirani (1993). Figure 6.2 displays a boxplot of the realized 5% expected shortfall test statistic ν^* from (6.2) for the UBS stock. The smooth expectile approach shows very small whiskers.

Figure 6.3 displays a similar boxplot of ν^* for realized 1% expected shortfall test statistic. Again the smooth expectile based approach shows very small whiskers but a bias seems to become obvious.

Table 6.1 gives a broader overview over a wider selection of assets. It displays bootstrapped 95% confidence intervals for ν_t^* given in (6.2) with $\theta = 0.05$. As before n_w denotes the moving window size, i.e. the estimation sample. Obviously the underlying assets could support larger estimation windows - especially for the extreme value theory and functional expectile based approaches - but the main interest of this survey are the small sample properties of the proposed estimators.

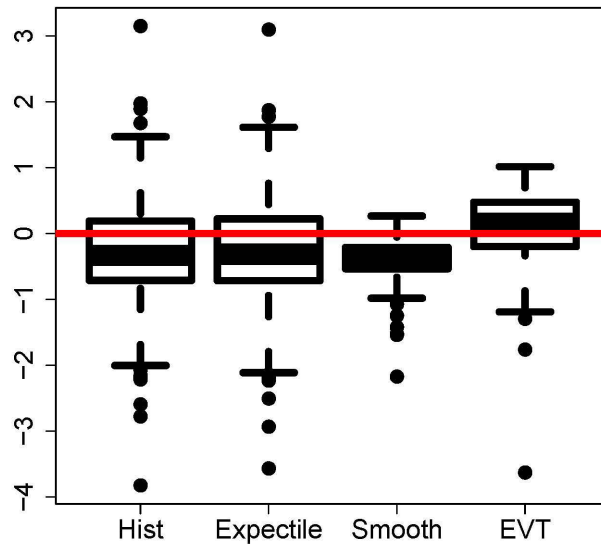


Figure 6.2: Conditional scaled differences (ν^*) of 5% expected shortfall estimators for the UBS stock returns. A moving window size n_w of 200 was employed.

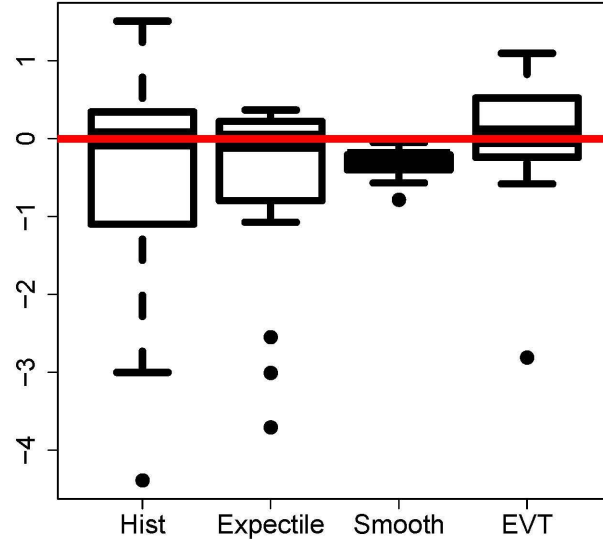


Figure 6.3: Conditional scaled differences (ν^*) of 1% expected shortfall estimators for the UBS stock returns. A moving window size n_w of 200 was employed.

 ESdynamicTest.R

n_w	Expectile	Functional	Historical	EVT
FTSE 100				
200	[-0.515, -0.240]	[-0.042, 0.077]	[-0.509, -0.217]	[-0.047, 0.128]
100	[-0.394, -0.136]	[-0.090, 0.014]	[-0.360, -0.032]	[-0.060, 0.168]
50	[-0.415, -0.148]	[-0.111, -0.027]	[-, -]	[-0.075, 0.123]
Cisco Stock				
200	[-1.031, -0.112]	[0.043, 0.114]	[-0.603, -0.236]	[-0.180, 0.106]
100	[-0.689, -0.284]	[-0.097, 0.013]	[-0.344, -0.096]	[-0.112, 0.090]
50	[-0.470, -0.205]	[-0.084, -0.010]	[-, -]	[-0.953, 41.040]
UBS Stock				
100	[-0.463, -0.176]	[-0.106, -0.008]	[-0.363, -0.042]	[-0.156, 0.171]
Brent Crude				
200	[-0.369, -0.185]	[-0.366, -0.316]	[-0.354, -0.156]	[-0.078, 0.050]
100	[-0.266, -0.180]	[-0.344, -0.277]	[-0.267, -0.127]	[-0.066, 0.028]

Table 6.1: 95% confidence intervals for ν^* with $\theta = 0.05$

 ESdynamicTest.R

Obviously the H_0 defined in 6.3 is least rejected for the extreme value theory approach. Nevertheless, the functional expectile approach shows very narrow confidence bands. This is confirmed by figure 6.4 that displays the scaled conditional difference between the realized quantile exceedances and a selected expected shortfall estimator as defined in 6.2 for the FTSE 100. The FTSE sample is displayed because it shows the challenges of expectile smoothing in this context most clearly. Estimation was carried out using a moving window of size 100. The functional expectile estimate captures the tail fairly well but is more sensitive to negative exceedances than to positive ones. This leads to it being outperformed (in a bias) sense by the extreme value based approach.

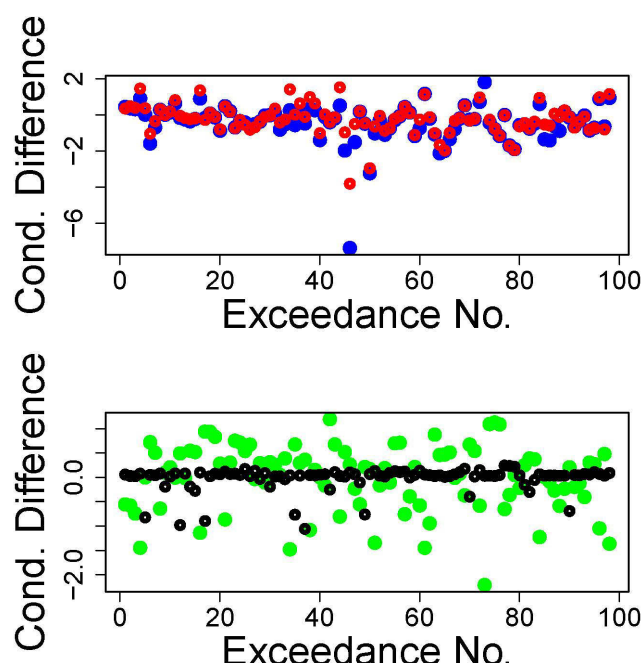


Figure 6.4: Test statistics ν_t^* (6.2) for testing the null of unbiasedness of four different expected shortfall estimators: historical (red), moving window expectile (blue), smooth expectile (black) and extreme value theory (green). Here we set $\theta=0.05$.

 ESdynamicTest.R

A look at table 6.1 reveals that the non-smooth rolling window expectile based approach and the historical approach suffer from a similar negative bias problem. This could indicate that the employed historical 100 day moving window quantile estimate is underestimating the true quantile. Further

evidence for this hypothesis is supplied by analyzing more extreme values of θ . Figure 6.5 displays the test statistic for $\theta = 1\%$. It becomes apparent that the historical and non-smooth expectile estimators suffer from the same asymmetric coverage problem as the smooth expectile approach. As the EVT based approach employs a pre-whitened loss series with the $\theta = 0.1$ quantile as threshold it is less affected by the inappropriate quantile estimate. First the returns were pre-whitened with a GARCH(1,1). This did not solve the challenge.

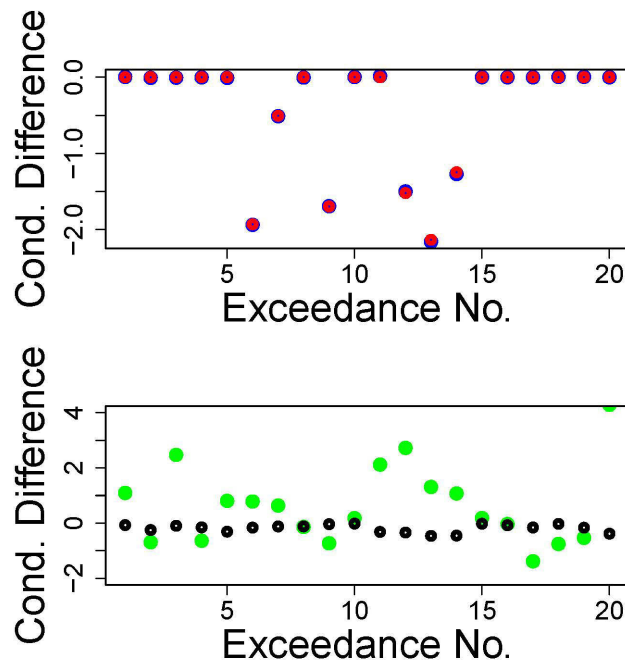


Figure 6.5: Conditional scaled differences of 1% expected shortfall estimators (Cisco stock)

 ESdynamicTest.R

Next, equation 4.16 was employed to apply an extreme value theory based quantile estimator with $\theta = 0.05$ and an adaptive 90% threshold u . Table 6.2 displays 95% confidence intervals for the expected shortfall test statistic in 6.2. The confidence intervals for the test statistics still indicates bias. Employing extreme value theory based quantile estimators did not pose a solution to the challenge of reducing the estimators bias. Nevertheless it should be stressed once again that the smooth expectile approach was able to generally deliver more accurate estimations than any other method in this study (see e.g. figure 6.4).

n_w	Expectile		Functional		Historical		EVT	
FTSE 100								
300	[-1.588,	-1.211]	[-0.244,	-0.107]	[N.D.,	N.D.]	[-0.023,	0.196]
150	[-1.084,	-0.691]	[-0.168,	-0.069]	[N.D.,	N.D.]	[-0.069,	0.084]
100	[-0.932,	-0.617]	[-0.147,	-0.072]	[N.D.,	N.D.]	[-0.047,	0.160]
Cisco Stock								
300	[-2.858,	-1.697]	[-0.285,	-0.085]	[N.D.,	N.D.]	[-0.197,	0.108]
150	[-1.640,	-1.052]	[-0.330,	-0.139]	[N.D.,	N.D.]	[-0.145,	0.064]
100	[-1.227,	-0.323]	[-0.226,	-0.069]	[N.D.,	N.D.]	[-0.150,	0.111]
UBS Stock								
150	[-1.355,	-0.833]	[-0.290,	-0.111]	[N.D.,	N.D.]	[-0.163,	0.133]

Table 6.2: ES test statistic confidence intervals (using EVT quantiles)

 ESdynamicTest.R

6.4.1 A Mincer-Zarnowitz regression

To find out whether the expectile approach delivers additional information to the extreme value theory and historical approaches a Mincer-Zarnowitz regression, Mincer and Zarnowitz (1969), for the conditional exceedances $r_t|_{r_t > q_{t,\theta}}$ is implemented. The condition $r_t > q_{t,\theta}$ is again evaluated using a GARCH(1,1) pre-whitened return series. This regression also allows for a combination of the expected shortfall estimations of all models included in the regression

$$r_t \left| \frac{r_t}{\sigma_t} > q_\theta = \beta_0 + \beta_1 \widehat{\text{ES}}_{\theta,t}^{\text{EXPECTILE}} + \beta_2 \widehat{\text{ES}}_{\theta,t}^{\text{HIST}} + \beta_3 \widehat{\text{ES}}_{\theta,t}^{\text{EVT}} + \beta_4 \widehat{\text{ES}}_{\theta,t}^{\text{SMOOTH}} \quad (6.4)$$

Table 6.3 displays 95% confidence intervals for the β_i , $i = \{0 \dots 4\}$ parameter from equation (6.4) with $\theta = 0.01$. Only time series yielding more than 50 valid exceedance returns $r_t|_{r_t > q_{t,\theta}}$ are listed. β_1 and β_2 have not found to be significant. Additional regressions were executed, excluding $\widehat{\text{ES}}_{\theta,t}^{\text{HIST}}$ or $\widehat{\text{ES}}_{\theta,t}^{\text{EXPECTILE}}$ to ensure that the results are not influenced by multicollinearity but the parameters were still insignificant. By far the strongest significance was found for the smooth expectile expected shortfall estimator.

The extreme value theory based estimator also showed slight significance, indicating that a combination of the two forecasts could yield an increased model fit.

For the BP return series the estimates for historical and moving window expectile estimation are in fact significant. But removing either one of them makes the other one insignificant, indicating multicollinearity.

	City Stock	Crude Oil	BP	BP
β_0	[0.029, 0.038]	[0.089, 0.237]	[-0.009, -0.002]	[-0.006, 0.000]
β_1	[-0.041, 0.229]	[-0.051, 0.262]	[0.114, 0.424]	[-0.056, 0.039]
β_2	[-0.229, 0.024]	[-0.320, 0.077]	[-0.466, -0.142]	
β_3	[-0.003, -0.000]	[-0.053, -0.001]	[-0.002, 0.000]	[-0.001, 0.000]
β_4	[0.900, 0.979]	[0.868, 0.957]	[0.830, 0.924]	[0.848, 0.945]

Table 6.3: 95% confidence intervals for the β parameters in regression (6.4)

 MincerZarno.R

6.5 Research Outlook: Larger samples

In the examined financial time series, for the smooth expectile approach, it was generally found that the portfolio risk was

- Underestimated if all parameters were estimated under a small $n_w \approx \in [140, 0)$
- Overestimated if all parameters (including e_τ) were estimated in a large sample ($n_w \gtrsim 1000$)
 - This is very likely reflecting the fact that τ in this setting is estimated nearly time static

Figure 6.6 displays the test statistic ν_t^* from equation 6.2 for the smooth expectile approach. The whole sample was used for expectile smoothing. q_θ and τ were calculated using a moving window of 600 observations, q_θ was computed extreme value theory based according to 4.16. The bootstrapped 95% confidence interval is $[-0.222, 0.308]$. Finding the optimal n_w for the

estimation of τ and q_θ could yield a possible extension of expectile based approaches. A simple solution would be to use the test statistic ν^* and numerically estimate n_w such that the bootstrapped 95% confidence interval is centered around 0.

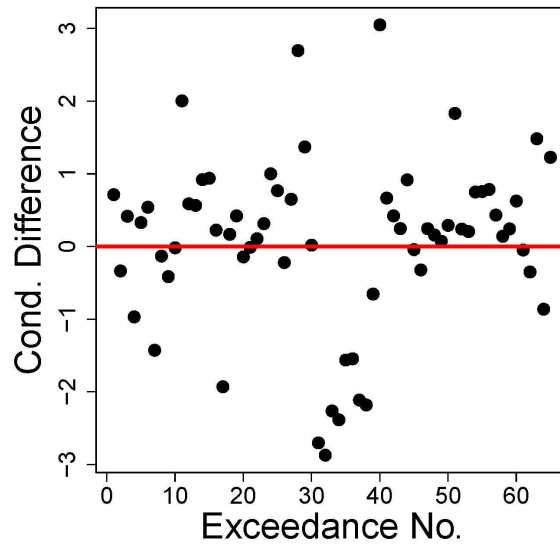


Figure 6.6: Conditional differences ν^* , for a smooth expectile estimator with $\theta = 5\%$ expected shortfall (Cisco stock). The smooth quantile function was calculated over the entire sample (6115 observations), whereas τ was calibrated using a moving window of size $n_w = 600$

Chapter 7

Summary

The usage of expected shortfall is very recommendable in risk management applications as it ensures that the risk is covered consistently, especially the tail risk structure. Its practical implementation however was seen to be troublesome for small samples.

Historical estimation needs to assume a high degree of stationarity and is very restrictive on the employed quantile estimators. It will break down as soon as the quantile estimator exceeds the largest observation. As the Basel committee recommends the usage of 1% risk measures this would in practice mean that the sample needs to include more than 100 reliable observations of the current underlying stochastic process to allow for this method.

Extreme value theory based approaches on the other hand were found to be unbiased in small samples but not very efficient as its ν^* statistic was very scattered for all considered assets.

It was found that expectile based approach did perform very promising. Especially the possibility to use smoothed quantile and expectile estimators did increase the prediction accuracy. Unfortunately it became obvious that the estimator is often underestimating the portfolio risk. This is most likely to be caused by an improper estimate of τ . This could possibly be fixed with more advanced q_τ estimators which would pose a natural extension of the approach. Additional extensions to achieve this are suggested below.

The proposed Mincer-Zarnowitz regression for expected shortfall did confirm the dominating fit of the functional expectile expected shortfall estimate.

Even though it is more sensitive to downside risk than to upside risk it was found to dominate risk measuring in terms of estimation efficiency against all other methods employed in this study, as even its worst conditional prediction error for ν_τ^* was always surpassed by the other models prediction errors, see e.g. figures 6.4 and 6.5.

7.1 Further extensions

The expectile based approach can be further extended. Apart from the above mentioned approaches there are further promising extensions which will especially increase the model's performance in small samples. The estimate for the expectile and its relation to the quantile can be obtained from a nonparametric expectile density estimation in the manner of Yao and Tong (1996).

Xiao and Koenker (2009) propose the estimation of quantiles with an underlying GARCH model structure. This approach could be altered accordingly to estimate the expectile.

Bibliography

- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking and Finance (Elsevier)* **26**(7).
- AG, E. C. (2007). Eurex risk based margining brochure.
- AG, E. C. (2013). Eurex clearing prisma brochure, http://www.eurexclearing.com/blob/clearing-en/51496-51498/294792/6/data/prisma_brochure.pdf.
- Aigner, D., T., A. and Poirier, D. (1976). On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function, *Journal of economic review* **17**(2).
- Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1998). Coherent measures of risk.
URL: <http://www.math.ethz.ch/delbaen/ftp/preprints/CoherentMF.pdf>
- Balkema, A. and de Haan, L. (1974). Residual life time at great age, *Annals of Probability* **2**.
- Bellini, F. and Bignozzi, V. (2013). Elicitable risk measures, *Working Paper*.
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2334746
- Bellini, F., Klar, B., Müller, A. and (2013), R. G. E. (2014). Generalized quantiles as risk measures, *Insurance: Mathematics and Economics* **54**.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**(3).
- Breckling, J. and Chambers, R. (1988). M-quantiles, *Biometrika* **75**(4).
- Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures, *Quantitative Finance* **10**(6): 593–606.

- Dana, R. A. . (2005). A representation result for concave schur concave functions, *Math Finance* **15**.
- Delbaen, F. (1998). Coherent risk measures on general probability spaces.
URL: *working paper*, <http://www.math.ethz.ch/delbaen>
- Dhaene, J., Denuit, M., Goovaerts, M. J. and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: theory, *Insurance: Mathematics & Economics* **31**(1).
- Diebold, F. X., Schuermann, T. and Stroughair, J. D. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management, *Journal of Risk Finance, The* **1**(2): 30–35.
- Efron, B. and Tibshirani, R. (eds) (1993). *An introduction to the bootstrap*, Chapman & Hall, New York.
- Embrechts, P., Frey, R. and McNeil, A. (2005a). Quantitative risk management, *Princeton Series in Finance*, Princeton **10**.
- Embrechts, P., Frey, R. and McNeil, A. J. (eds) (2005b). *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton university press.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (eds) (1997). *Modelling Extremal Events: For Insurance and Finance*, Springer.
- Embrechts, P., Lambrigger, D. and Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples, *Extremes* **12**.
- Emmer, S., Kratz, M. and Tasche, D. (2013). What is the best risk measure in practice? a comparison of standard measures.
URL: <http://arxiv.org/abs/1312.1645>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* **50**(4).
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints, *Finance and Stochastics* **6**. Springer.
- Föllmer, H. and Schied, A. (2008). Convex and coherent risk measures.
URL: <http://www.alexschied.de/Encyclopedia6.pdf>
- Föllmer, H. and Schied, A. (eds) (2004). *Stochastic Finance: An Introduction in Discrete Time*, number 2 in 27, Walter de Gruyter & Co., Berlin, de Gruyter Studies in Mathematics.

- Franke, J., Härdle, W. K. and Hafner, C. M. (eds) (2010). *Statistics of Financial Markets: An Introduction*, Vol. 3, Springer-Verlag Berlin Heidelberg.
- Frittelli, M. and Rosazza Gianin, E. (2005). Law-invariant convex risk measures, *Advances in Mathematical Economics* **7**.
- Gneiting, T. (2011). Making and evaluating point forecasts, *Journal of the American Statistical Association* **106**(494): 746–762.
- Heath, D. (2000). Back to the future (plenary lecture), *First World Congress of the Bachelier Finance Society* **1**.
- Huber, P. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics* **35**(1).
- Huber, P. (ed.) (1981). *Robust statistics*, Wiley, New York.
- Huber, P. and Ronchetti, E. (eds) (2009). *Robust statistics*, number 2, Wiley, New York.
- Jones, M. (1994a). Expectiles and m-quantiles are quantiles, *Statistics and probability letters* **20**.
- Jones, M. C. (1994b). Expectiles and m-quantiles are quantiles, *Statistics & Probability Letters* **20**(2): 149–153.
- Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the fatou property, *Advances in Mathematical Economics* **9**.
- Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica* **46**(1).
- Koenker, R. (ed.) (2005). *Quantile Regression*, Vol. 38, Cambridge University Press, Econometric society monographs.
- Kunze, M. (2003). *Verteilungsinvariante konvexe Risikomasse*, Master's thesis, Humboldt Universität zu Berlin.
- Kusuoka, S. (2001). *On law invariant coherent risk measures*, Vol. 3, Advances in Mathematical Economics.
- McNeil, A. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach, *Journal of Empirical Finance* **7**.

- Mincer, J. and Zarnowitz, V. (1969). The evaluation of economic forecasts, *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER .
- Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the basel committee, *Bank of Italy, Temi di discussione (Economic working papers)* **517**.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing, *Econometrica* **55**(4).
- on banking supervision, B. C. (2014). Fundamental review of the trading book: A revised market risk framework.
URL: <http://www.bis.org/publ/bcbs265.pdf>
- Pflug, G. and Wozabal, D. (2007). Ambiguity in portfolio selection, *Quantitative Finance* **7**.
- Pichler, A. ((forthcoming)). Insurance pricing under ambiguity, *European Actuarial Journal* .
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics* **3**.
- RiskMetricsTM* - *Technical Document* (1994). *Technical report*, J. P. Morgan and Reuters.
URL: http://pascal.iseg.utl.pt/~afonso/eif/rm/TD4ePt_2.pdf
- Rockafeller, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* **26**.
- Schall, R. (1991). Estimation in generalized linear models with random effects, *Biometrika* **78**.
- Schnabel, S. K. and Eilers, P. H. (2009). Optimal expectile smoothing, *Computational Statistics & Data Analysis* **53**(12): 4168–4177.
- Smith, R. L. (1987). Estimating tails of probability distributions, *Annals of Statistics* **15**.
- Sobotka, F., Schnabel, S., Schulze Waltrup, L., Eilers, P., Kneib, T. and Kauermann, G. (2011). expectreg: Expectile and quantile regression, *R package version 0.25* .

- Stahl, G., Zheng, J., Kiesel, R., and Rülicke, R. (2012). Conceptualizing robustness in risk management.
URL: *SSRN*: <http://ssrn.com/abstract=2065723>
- Tasche, D. (2002). Expected shortfall and beyond.
URL: <http://arxiv.org/pdf/cond-mat/0203558.pdf>
- Tasche, D. (2013). Expected shortfall is not elicitable, so what?
URL: https://workspace.imperial.ac.uk/mathfn/Public/Seminars%202013-2014/Tasche_November2013_Slides.pdf
- Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics* **6**.
- Xiao, Z. and Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models, *Journal of the American Statistical Association* **104**(488).
- Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach, *Journal of Nonparametric Statistics* **6**(2-3): 273–292.
- Ziegel, J. F. (2013). Coherence and elicibility, *ArXiv preprint*, *arXiv:1303.1690v2*.

Anhang A

$$(1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau) f(u) du =$$

$$\tau \int_{e_\tau}^{\infty} (e_\tau - u) f(u) du + \tau \int_{-\infty}^{e_\tau} e_\tau f(u) du - \tau \int_{-\infty}^{e_\tau} e_\tau f(u) du$$

$$(1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau) f(u) du =$$

$$\tau \int_{-\infty}^{\infty} e_\tau f(u) du - \tau \int_{e_\tau}^{\infty} u f(u) du - \tau \int_{-\infty}^{e_\tau} e_\tau f(u) du$$

$$(1 - \tau) \int_{-\infty}^{e_\tau} (u - e_\tau) f(u) du + \tau \int_{e_\tau}^{\infty} u f(u) du =$$

$$\tau \int_{-\infty}^{\infty} e_\tau f(u) du - \tau \int_{-\infty}^{e_\tau} e_\tau f(u) du$$

$$(1 - \tau) \int_{-\infty}^{e_\tau} u f(u) du + \tau \int_{e_\tau}^{\infty} u f(u) du =$$

$$(1 - \tau) \int_{-\infty}^{e_\tau} e_\tau f(u) du + \tau \int_{-\infty}^{\infty} e_\tau f(u) du - \tau \int_{-\infty}^{e_\tau} e_\tau f(u) du$$

$$(1 - \tau) \int_{-\infty}^{e_\tau} u f(u) du + \tau \int_{e_\tau}^{\infty} u f(u) du =$$

$$e_\tau \left\{ (1 - \tau) \int_{-\infty}^{e_\tau} f(u) du + \tau \int_{-\infty}^{\infty} f(u) du - \tau \int_{-\infty}^{e_\tau} f(u) du \right\}$$

$$\frac{(1 - \tau) \int_{-\infty}^{e_\tau} u f(u) du + \tau \int_{e_\tau}^{\infty} u f(u) du}{(1 - \tau) \int_{-\infty}^{e_\tau} f(u) du + \tau \int_{-\infty}^{\infty} f(u) du - \tau \int_{-\infty}^{e_\tau} f(u) du} = e_\tau$$

$$\frac{(1 - \tau) \int_{-\infty}^{e_\tau} u f(u) du + \mu + \tau \int_{-\infty}^{e_\tau} u f(u) du}{(1 - \tau) F(e_\tau) + \tau(1 - F(e_\tau))} = e_\tau$$

□

Declaration of Authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Student's signature: _____

Name: PHILIPP GSCHÖPF

Date of submission: July 18, 2014